

PRODUCT INTRODUCTION

Argus Catalog

제품 소개

데이터 · ML 모델 · API · AI 에이전트 — 흩어져 있던 자산을 하나의 카탈로그에서 찾고, 관리하고, 거버넌스합니다.

DataHub 스타일 데이터 카탈로그 + Unity Catalog OSS 호환 모델 레지스트리를 하나로 묶은 오픈소스 메타데이터 플랫폼

제품 유형

오픈소스 메타데이터 플랫폼 — Apache 2.0 · 에어갭(폐쇄망) 지원

관리 자산

데이터셋 · ML 모델 · API · AI Agent (4종 통합)

기술 스택

FastAPI · Next.js · PostgreSQL(pgvector) · MinIO/S3

문서 기준일

2026-07-17

Data Dynamics Inc

Argus Catalog

무엇

데이터셋·스키마·리니지·품질·표준·용어집의 데이터 카탈로그와 MLflow·OCI 호환 모델 레지스트리, 그리고 API·AI 에이전트 카탈로그까지 — 전사 자산 메타데이터를 단일 플랫폼에서 관리합니다.

왜

자산이 늘수록 "어디에 무엇이 있는지"가 병목이 됩니다. 카탈로그·품질·모델 레지스트리를 도구별로 따로 설치·연결하면 운영 부담이 커지고, 폐쇄망에서는 아예 선택지가 없는 경우가 많습니다.

얻는 것

임베딩 기반 하이브리드 검색으로 자산을 즉시 발견하고, 컬럼 수준 리니지·품질 점수·표준 준수율로 신뢰도를 확인하며, 에어갭 모델 반입과 카탈로그 페더레이션으로 폐쇄망·멀티 사이트까지 커버합니다.

4종

관리 자산 — 데이터셋·모델·API·에이전트

11종

데이터 소스 메타데이터 동기화

10종

데이터 품질 검증 규칙

Apache 2.0

오픈소스 · 에어갭 지원

※ 여러 도구를 따로 설치·연결할 필요 없이 단일 저장소로 제공 — 필수 의존성은 PostgreSQL(pgvector)과 오브젝트 스토리지 둘뿐입니다.

목차

Part 1	개요와 아키텍처	제품 포지셔닝(자산 4종), 플랫폼 아키텍처, 기술 스택과 화면 구성	04 - 06
Part 2	데이터 카탈로그	데이터셋·메타데이터 동기화, 리니지·ERD, 데이터 표준·용어집, 시맨틱 검색	07 - 10
Part 3	데이터 품질	소스 직접 프로파일링과 규칙 검증, 품질 점수·리니지 전파	11 - 12
Part 4	거버넌스 확장	API 카탈로그, AI Agent 카탈로그, 변경관리·알림·권한	13 - 15
Part 5	모델 레지스트리 · AI	MLflow·OCI 모델 레지스트리, 에어갭 모델 반입, AI 메타데이터·어시스턴트	16 - 18
Part 6	페더레이션과 차별점	카탈로그 페더레이션, 확장 생태계, 보안·운영, 차별점 요약	19 - 22
Part 7	Data Dynamics 서비스	도입 시나리오 · 제공 서비스 · 문의 및 Next Steps	23 - 24

제품 포지셔닝 — 자산 4종을 하나의 카탈로그에서

데이터만 다루는 카탈로그가 아닙니다. 데이터셋과 동급의 1급(first-class) 엔티티로 모델·API·AI 에이전트를 등록하고, 소유자·태그·리니지·검색·거버넌스를 공통으로 적용합니다.

데이터셋

11종 소스 메타데이터 동기화, 컬럼 수준 리니지·ERD, 데이터 표준·용어집, 품질 점수 — DataHub 스타일 데이터 카탈로그.

ML 모델

MLflow·OCI 호환 레지스트리 — 버전/Stage 관리, 메트릭 비교, 모델 카드, OCI 모델 허브와 에어갭 반입.

API

OpenAPI 스펙 등록·문서화·버전 diff·린트 — URN 기반 메타데이터 거버넌스로 API 자산화.

AI Agent

에이전트 등록·도구/MCP·실행정책·가드레일(DLP·HITL·예산)·평가·미터링 — "그림자 AI" 방지.

핵심 차별점 메타데이터 공통 기반(소유자·태그·리니지·댓글·검색·LLM 자동 설명) 위에 자산별 고유 속성만 있는 구조 — 자산이 늘어도 거버넌스 방식은 하나입니다.

플랫폼 아키텍처

Next.js 웹 UI에서 FastAPI 백엔드(:4600)로 — 카탈로그·품질·검색·페더레이션·모델 허브가 PostgreSQL(pgvector)과 오브젝트 스토리지 위에서 동작합니다.



※ 독립 모듈: agent/(AI 메타데이터·어시스턴트, 로컬 LLM 연동) · quality/(pandas/PySpark 외부 품질 배치) · sdk/(argus-model CLI) — 필요할 때만 붙입니다.

기술 스택과 화면 구성

레이어	구성
Backend	Python 3.11+ · FastAPI · SQLAlchemy · Uvicorn
데이터	PostgreSQL(pgvector) · MinIO/S3 호환 오브젝트 스토리지
Frontend	Next.js · React · TypeScript · pnpm
모델 레지스트리	MLflow 호환 · OCI 매니페스트
임베딩	Local(sentence-transformers · 오프라인) / OpenAI / Ollama
AI (LLM)	OpenAI · Anthropic · Ollama (OpenAI 호환 엔드포인트)

대시보드 화면 (주요 메뉴)

통합 검색	데이터셋	데이터소스
택소노미	태그	용어집
데이터 표준	모델	OCI 허브
MLflow 파일	AI Agent	API
페더레이션	변경관리	내 승인
알림	사용자/권한	설정

※ 검색부터 승인 워크플로까지 단일 웹 UI — 오픈소스(Apache 2.0)라 라이선스 비용 없이 시작할 수 있습니다.

⚙ SemVer 버전 규약 · CHANGELOG · CI — 오픈소스 프로젝트 운영 체계를 갖추고 Data Dynamics가 후원·기여합니다.

데이터셋 관리와 메타데이터 동기화

11종 데이터 소스에서 스키마·테이블 메타데이터를 동기화하고, 소유자·태그·택소노미로 조직화합니다.

동기화 지원 데이터 소스 (11종)

Hive · Impala · Trino · StarRocks · MySQL · PostgreSQL · Kudu · HDFS · S3 · Kafka
· MongoDB · Elasticsearch 등

빅데이터(Hadoop 계열)와 RDBMS·NoSQL·스트리밍·스토리지를 함께 커버 — 국내 엔터프라이즈 데이터 플랫폼 구성에 맞춘 라인업

- **소유자 3역할** — Technical Owner · Business Owner · Data Steward — 기술/비즈니스/관리 책임을 분리해 지정.
- **URN 기반 식별** — {namespace}.{name} 체계의 URN으로 자산을 전사 고유하게 식별.
- **확장 메타데이터** — 택소노미(트리 분류) · 태그 · 댓글 · 시스템/조직 토폴로지 연계.

동기화 방식

- 소스별 커넥터가 스키마·테이블·컬럼 메타데이터 수집
- 수동 동기화 + 스케줄 동기화(인제스천 스케줄)
- 스키마 변경 감지 → 영향 분석·알림 연계
- 데이터셋별 PySpark 접근 코드 스니펫 자동 제공

※ 메타데이터만 수집합니다 — 원본 데이터는 소스에 그대로 두므로 데이터 이동·복제 리스크가 없습니다.

리니지 & ERD — 데이터 흐름을 컬럼 수준까지

"이 컬럼은 어디서 왔고, 바꾸면 무엇이 깨지는가"에 답합니다. 이기종 시스템을 가로지르는 컬럼 수준 리니지와 DDL 파싱 기반 ER 다이어그램을 제공합니다.

컬럼 수준 리니지

- **이기종 추적** — Hive → Trino → StarRocks처럼 시스템 경계를 넘는 데이터 흐름을 컬럼 단위로 추적.
- **쿼리 로그 자동 수집** — hive-query-hook · trino-query-listener · starrocks-audit-plugin · impala-query-agent 확장이 실행 쿼리를 수집, SQL 파싱(sqlglot)으로 리니지 자동 구축.
- **영향 분석** — 스키마 변경 시 다운스트림 영향 분석 → 알림으로 통지.

ERD (ER 다이어그램)

- **DDL 파싱 기반** — 테이블 DDL에서 관계를 파싱해 ER 다이어그램 자동 생성.
- **관계 분석기** — argus-relationship-analyzer — 조인 패턴·데이터 기반 관계 추론 확장.
- **Mermaid 내보내기** — 다이어그램을 Mermaid로 내보내 문서·위키에 그대로 사용.

※ 페더레이션 환경에서는 인스턴스 경계를 넘는 cross-instance 리니지까지 이어집니다(19장).

데이터 표준과 용어집 — 한국어 환경에 맞춘 표준화

공공·금융의 데이터 표준화 요구를 겨냥한 축입니다. 멀티 표준 사전과 형태소 분석 기반 용어 자동 생성으로 표준화 작업량을 줄입니다.

멀티 표준 사전

표준 단어·용어·도메인 사전을 복수로 운영 — 기관 표준과 전사 표준을 함께 관리합니다.

형태소 분석 용어 자동 생성

컬럼·테이블명을 형태소 분석해 표준 용어 후보를 자동 생성 — 한국어 명명 체계에 특화된 기능입니다.

준수율 측정

데이터셋·컬럼의 표준 준수율을 측정해 표준화 진행도를 숫자로 관리합니다.

트리 분류 용어집

비즈니스 용어를 트리 구조로 분류하고 데이터셋·API 등 자산에 연결 — 용어 기반 탐색을 지원합니다.

※ 표준·용어집은 검색과 연결됩니다 — 용어로 찾고, 표준 준수율로 신뢰도를 판단하는 흐름이 한 화면에서 이어집니다.

시맨틱 검색 — 키워드 + 의미 기반 하이브리드

정확한 이름을 몰라도 찾을 수 있어야 카탈로그입니다. pgvector 임베딩 검색과 키워드 검색을 결합해, 자산 4종을 하나의 검색창에서 찾습니다.

1. 질의

"고객 결제 이력" — 자연어로 검색

2. 하이브리드 매칭

키워드 매칭 + pgvector 임베딩 유사도 결합

3. 자산 통합 결과

데이터셋 · 모델 · API · 에이전트를 한 결과에서

4. 필터·탐색

태그 · 소유자 · 소스 · 품질 등급으로 좁히기

- **임베딩 프로바이더 선택** — Local(sentence-transformers — 완전 오프라인) / OpenAI(text-embedding-3 등) / Ollama(nomic-embed-text 등) — 폐쇄망에서는 로컬 임베딩으로 외부 연결 없이 시맨틱 검색이 동작합니다.
- **메타데이터가 검색 품질을 만든다** — LLM 자동 생성 설명·태그(18장)가 임베딩 대상에 포함되어, 카탈로그를 채울수록 검색이 좋아지는 구조입니다.
- **페더레이션 통합 검색** — 연합된 여러 인스턴스를 하나의 검색으로 — HARVEST 미러는 로컬 임베딩 모델로 재임베딩해 검색 품질을 유지합니다(19장).

프로파일링과 규칙 검증 — 소스에 직접

데이터를 옮겨와서 검사하는 방식이 아닙니다. 소스 DB에 직접 프로파일링·검증을 수행하고, 대규모 배치는 외부 품질 잡(pandas/PySpark)으로 분리 실행합니다.

프로파일링

- **소스 직접 프로파일링** — 행 수·NULL 비율·고유값·min/max·최빈값(top values)까지 — 소스 DB에서 직접 수집.
- **컬럼 통계 이력** — 프로파일 결과를 시점별로 축적해 분포 변화를 추적.
- **Impala 프로파일링** — 빅데이터 환경(Impala) 전용 프로파일 경로 제공.

규칙 검증 — 10종

- **기본 규칙 8종** — NOT NULL · 고유성 · 범위 · 패턴 등 표준 검증 규칙.
- **CUSTOM_SQL · CUSTOM_PYTHON** — SQL·Python으로 도메인 규칙을 자유롭게 정의 — 표준 규칙으로 안 되는 검증까지.
- **외부 품질 배치** — quality/ 모듈 — pandas/PySpark 잡으로 대용량 검증을 카탈로그 밖에서 실행하고 결과만 회수.

※ 검증 결과는 데이터셋 품질 점수로 집계되고, 실패는 알림 트리거·Webhook으로 즉시 통지됩니다(12장).

품질 점수와 리니지 전파 — 신뢰도를 한눈에

품질은 등급(GOOD/WARN/BAD)으로 자동 집계되고, 리니지를 따라 전파됩니다 — 업스트림이 나쁘면 다운스트림 사용자에게 경고가 표시됩니다.

1. 검증 실행

규칙 검증·프로파일링 결과 수집

2. 점수 산정

GOOD / WARN / BAD 자동 동기화 · 추세 추적

3. 리니지 전파

업스트림 품질 문제를 다운스트림에 경고로 표시

4. 통지

품질 실패 트리거 → 알림 · Webhook

왜 중요한가

카탈로그의 진짜 질문은 "찾을 수 있는가"가 아니라 "믿고 써도 되는가"입니다. 품질 점수가 데이터셋 카드에 붙고, 나쁜 품질이 리니지를 따라 경고로 전파되면 — 분석가는 잘못된 데이터로 보고서를 만들기 전에 위험을 알게 됩니다. 품질 이력·추세까지 남아 개선 여부를 추적할 수 있습니다.

※ 품질 등급·추세는 검색 필터와 데이터셋 상세에 노출됩니다 — 거버넌스 지표가 아니라 사용자의 일상 판단 기준이 됩니다.

API 카탈로그 — API를 자산으로 등록·문서화·거버넌스

SwaggerHub·Postman·Backstage 등 시장 솔루션을 벤치마크해 설계한 API 1급 엔티티 — 스펙 등록부터 문서·버전·거버넌스까지 카탈로그 안에서 처리합니다.

스펙 등록·동기화

OpenAPI 2.0/3.x를 파일·URL·붙여넣기로 등록, 원격 URL 주기 동기화로 스펙 드리프트 감지 — 유효성 검증과 오류 위치 표시.

문서화·버전 diff

스펙 파싱 기반 문서 렌더, 버전 간 diff로 변경점 추적, 스타일 린트 — 소비자가 변경 영향을 미리 파악.

탐색·거버넌스

소유자·부서·시스템 연계, 도메인·태그·용어집 연결, URN 기반 식별 — 데이터셋과 같은 방식으로 검색·관리.

데이터 자산과의 연결

API가 어떤 데이터셋·시스템 위에 서 있는지 토폴로지로 연결 — "이 테이블 바꾸면 어느 API가 깨지나"에 답합니다.

※ AsyncAPI · gRPC/Protobuf · GraphQL SDL 등 추가 포맷은 설계 로드맵에 포함되어 있습니다.

AI Agent 카탈로그 — "그림자 AI"를 거버넌스로

부서마다 만들어지는 AI 에이전트를 카탈로그에 등록해 누가·무엇을·어떤 권한으로 실행하는지 관리합니다. 데이터셋·모델과 동급의 1급 엔티티입니다.

등록 · 식별

agent_id·URN·버전(SemVer)·상태(draft→active→retired) — 소유자·부서·도메인과 함께 전사 고유 관리.

도구 · MCP

에이전트가 사용하는 도구와 MCP 서버를 메타데이터로 명세 — 무엇에 접근하는지 투명하게.

가드레일 · 실행정책

DLP·네트워크 화이트리스트·예산 한도·HITL(사람 개입) 등 보안 메타데이터를 1차 릴리스부터 필수 스키마로 — "그림자 액션" 방지.

평가 · 미터링

에이전트 평가 지표와 사용량 미터링 — 비용과 품질을 함께 추적. 리니지·버전 관리 공통 기능 재사용.

※ AI 도입이 확산될수록 "어떤 에이전트가 어떤 데이터·도구에 닿는가"가 보안 이슈가 됩니다 — 데이터 거버넌스와 같은 틀로 AI 거버넌스를 시작하세요.

변경관리 · 알림 · 권한 — 거버넌스를 워크플로우로

메타데이터 변경을 승인 워크플로우로 통제하고, 스키마 변경·품질 실패를 알림으로 전파하며, RBAC로 접근을 관리합니다.

변경관리 (Temporal)

메타데이터 변경을 제출 → 승인 → 적용하는 워크플로우 — Temporal 엔진 기반, "내 승인" 화면에서 처리. 그면(submit 비활성) 단순 모드로 운영 가능.

알림

스키마 변경 영향 분석 통지, 품질 실패 트리거, Webhook 연동 — 카탈로그가 변화를 먼저 알려줍니다.

권한 (RBAC)

역할 기반 접근 제어와 사용자 관리 — 로컬 인증 기본, Keycloak OIDC/SSO·LDAP 동기화 선택.

자격증명 보호

소스 접속 정보는 ARGUS_SECRET_KEY로 암호화 저장 — JWT 서명키 (ARGUS_JWT_SECRET)와 함께 운영 보안 체크리스트 제공.

※ 규제 산업의 "메타데이터도 통제 대상" 요건에 대응 — 변경 이력과 승인 기록이 남습니다.

ML 모델 레지스트리 — MLflow · OCI 호환

Unity Catalog OSS 호환 모델 레지스트리 — 학습 산출물을 버전·Stage로 관리하고, 메트릭을 비교하며, 모델 카드로 문서화합니다.

- **버전 · Stage 관리** — 모델 버전마다 STAGING/PRODUCTION Stage 승격 — 어떤 버전이 운영 중인지 항상 명확합니다.
- **MLflow 연동** — 기존 MLflow 실험·메트릭을 연동해 버전 간 메트릭 비교 — 학습 워크플로를 바꾸지 않고 거버넌스만 엮습니다.
- **모델 카드** — 용도·한계·학습 데이터를 문서화 — 데이터셋과 같은 소유자·태그·검색 체계.
- **아티팩트 저장** — S3/MinIO 호환 오브젝트 스토리지에 모델 파일 관리 — MLflow 파일 브라우저 내장.

데이터 ↔ 모델을 잇는 카탈로그

모델이 어떤 데이터셋으로 학습되었는지, 그 데이터의 품질·리니지는 어떤지 — 데이터 카탈로그와 모델 레지스트리가 한 플랫폼이라 이 연결이 자연스럽습니다.

별도 도구 2~3개(카탈로그 + 레지스트리 + 허브)를 운영·연동하는 비용이 사라집니다.

※ OCI 매니페스트 표준을 따르므로 모델 배포 파이프라인(컨테이너 도구체인)과 자연스럽게 연결됩니다.

OCI 모델 허브와 에어갭 반입 — 폐쇄망에 모델 공급

인터넷이 없는 사내망에서 HuggingFace 모델을 쓰려면 반입 체계가 필요합니다. argus-model CLI/SDK가 그 파이프라인을 제공합니다.

1. 외부망에서 패키징

argus-model CLI로 HuggingFace 등 외부 모델을 OCI 형식으로 팩

2. 반입

승인 절차를 거쳐 폐쇄망으로 매체 반입

3. Airgap 임포트

OCI Model Hub로 임포트 — 매니페스트·레이어 검증

4. 탐색 · 사용

HuggingFace 스타일 모델 브라우저에서 검색·다운로드

- **OCI Model Hub** — 사내 HuggingFace처럼 동작하는 모델 허브 — OCI 매니페스트 기반 저장·버전·태그, S3/MinIO 백엔드.
- **Python SDK + CLI** — argus-model 명령으로 패키징·업로드·다운로드 자동화 — CI/CD와 배치에서 재사용.
- **거버넌스 연결** — 반입된 모델도 카탈로그 자산으로 등록 — 누가 반입했고 어디에 쓰이는지 추적됩니다.

※ Argus RAG Studio 등 사내 AI 플랫폼의 모델 공급 체계로도 활용됩니다 — 폐쇄망 AI 스택의 공통 기반.

AI 기능 — 메타데이터 자동 생성과 어시스턴트

카탈로그의 고질병은 "비어 있는 설명란"입니다. LLM이 초안을 만들고 사람이 승인하는 흐름으로 카탈로그를 빠르게 채우고, tool-use 어시스턴트로 대화형 탐색까지 제공합니다.

AI 메타데이터 생성

- **설명 · 태그 자동 생성** — 스키마 · 샘플을 근거로 데이터셋 설명과 태그를 LLM이 제안.
- **PII 감지** — 개인정보 의심 컬럼을 감지해 표시 — 거버넌스 후속 조치로 연결.
- **미리보기 후 승인** — AI 제안은 항상 사람이 검토 · 승인 후 반영 — 자동 오염 방지.

AI 어시스턴트 (tool-use 채팅)

- **도구 호출 기반** — 카탈로그 검색 · 스키마 · ERD · 품질 · 리니지 · SQL 도구를 호출해 실행
이터 근거로 답변 — 환각이 아니라 조회 결과로 말합니다.
- **프로바이더 선택** — OpenAI · Anthropic · Ollama(OpenAI 호환) — 폐쇄망은 로컬
LLM(Ollama 등)으로.
- **독립 에이전트 모듈** — agent/ 를 별도 프로세스로 배포 — 카탈로그 본체와 분리 확장.

※ LLM은 선택 의존성입니다 — 없어도 카탈로그 핵심 기능은 전부 동작하고, 붙이면 채우는 속도가 달라집니다.

카탈로그 페더레이션 — 여러 인스턴스를 하나로

본사·계열사·망 분리 구역마다 카탈로그가 따로 있어도, 연합하면 하나처럼 검색·탐색합니다. 에어갭 환경을 고려한 HARVEST 미러 방식이 특징입니다.

연합 모드 3종

LIVE(실시간 질의) · HARVEST(주기 수집 미러) · HYBRID — 네트워크 상황에 맞게 인스턴스 별 선택.

통합 검색·탐색

연합 인스턴스 전체를 하나의 검색으로 — 트리+그리드 탐색, cross-instance 리니지 연결.

에어갭 친화 HARVEST

미러를 로컬로 복제하고 로컬 임베딩 모델로 재임베딩, 샘플 데이터 미러링 — 망이 끊겨도 탐색이 동작합니다.

거버넌스·안정성

visibility 정책으로 노출 범위 통제, 서비스 토큰 인증, circuit breaker·관측성 — 로컬 승격(import)으로 미러를 정식 자산으로 복사(태그 매칭·페더레이션 마커).

※ 중앙 집중형 단일 카탈로그를 강제하지 않습니다 — 조직·망 구조를 그대로 두고 메타데이터만 연합합니다.

확장 생태계 — 플러그인과 수집기

extensions/ 의 독립 구성요소들이 쿼리 수집·메타데이터 동기화·관계 분석을 담당합니다. 플러그인 구조라 필요한 것만 골라 배포합니다.

구성요소	역할	구분
metadata-sync	데이터 소스 메타데이터 주기 동기화 배치	수집
hive-query-hook · trino-query-listener starrocks-audit-plugin · impala-query-agent	각 엔진의 실행 쿼리 로그 수집 → 리니지 자동 구축의 원천	수집
argus-sql-parse · sqlglot-impala	SQL 파싱 — 쿼리에서 테이블·컬럼 관계 추출	분석
source-analyzer · relationship-analyzer	소스 구조 분석 · 데이터 기반 테이블 관계 추론(ERD 보강)	분석
query-service	카탈로그에서 소스 데이터 조회·샘플 질의 서비스	서비스
ldap-user-sync	LDAP 사용자 주기 동기화	운영
argus-plugin-acme	커스텀 플러그인 개발 예제 — 자체 확장의 출발점	예제

※ 테스트 스위트가 Hive·Trino·StarRocks·PostgreSQL 등 플랫폼별 통합 테스트를 포함 — 지원 소스가 문서 속 약속이 아니라 검증 대상입니다.

보안 · 운영

페쇄망 기본 설계에, 오픈소스 프로젝트다운 투명한 운영 체계를 갖췄습니다.

인증

로컬 JWT 기본(외부 의존 없음), Keycloak OIDC/SSO·LDAP 동기화 선택 — 페쇄망 친화.

접근 제어

RBAC · 사용자/권한 관리 · 페더레이션 visibility 정책과 서비스 토큰.

버전 · 릴리스

SemVer 규약, 루트 VERSION 단일 기준, CHANGELOG · CI(GitHub Actions) · SECURITY.md.

암호화

소스 자격증명 암호화 저장(ARGUS_SECRET_KEY), JWT 서명키 분리(ARGUS_JWT_SECRET).

운영 보안 가이드

개발 기본값 → 운영 전환 체크리스트를 문서로 제공 — 기본 크리덴셜 교체 항목 명시.

라이선스

Apache 2.0 오픈소스 — 벤더 종속 없이 시작, Data Dynamics가 후원·기여 및 상용 지원.

※ 메타데이터 플랫폼이므로 원본 데이터는 소스에 남습니다 — 카탈로그 서버가 다뤄야 할 민감 정보 표면적 자체가 작습니다.

차별점 요약 — 왜 Argus Catalog인가

단일 저장소, 가벼운 발

카탈로그·품질·모델 레지스트리·API/에이전트 카탈로그를 도구별로 설치·연결하는 대신 하나로 — 필수 인프라는 PostgreSQL과 오브젝트 스토리지뿐이라 도입 장벽이 낮습니다.

에어갭 네이티브

로컬 임베딩·로컬 LLM·OCI 모델 반입·HARVEST 페더레이션 — "인터넷 없이 전부 동작"이 설계 전제입니다.

리니지의 자동화

쿼리 로그 훅(Hive/Trino/StarRocks/Impala) + SQL 파싱으로 리니지를 자동 구축 — 수작업 등록에 의존하지 않습니다.

자산 4종 통합 거버넌스

데이터에서 멈추지 않고 모델·API·AI 에이전트까지 같은 틀로 — AI 시대의 거버넌스 범위를 처음부터 반영했습니다.

한국 환경 특화

형태소 분석 용어 자동 생성, 공공·금융형 데이터 표준 관리, Hive·Impala·Trino·StarRocks·Kudu 등 국내 빅데이터 스택 커넥터.

오픈소스 + 상용 지원

Apache 2.0로 시작하고, Data Dynamics의 구축·기술지원으로 운영 리스크를 줄입니다.

※ 같은 Argus 제품군의 RAG Studio와 결합하면 "자산을 찾는 카탈로그"와 "자산으로 답하는 RAG"가 하나의 스택이 됩니다.

도입 시나리오와 제공 서비스

전사 데이터 거버넌스

표준·용어집·품질·리니지를 갖춘 데이터 카탈로그 구축 — 데이터 표준화 사업·거버넌스 체계 수립에.

폐쇄망 AI 자산 관리

에어갭 모델 반입 + 모델 레지스트리 + AI 에이전트 카탈로그 — 망분리 환경의 AI 도입 기반.

멀티 사이트 카탈로그 연합

본사·계열사·망 구역별 인스턴스를 페더레이션으로 통합 검색 — 조직 구조를 바꾸지 않는 통합.

구축 · 컨설팅

PoC부터 전사 구축까지.

- ✓ 카탈로그·표준 체계 설계
- ✓ 소스 연동·리니지 구축
- ✓ 페더레이션·에어갭 설계

기술지원

오픈소스 운영의 안전망.

- ✓ 설치·업그레이드 지원
- ✓ 성능·운영 튜닝
- ✓ 장애 대응·에스컬레이션

교육 · 커스터마이징

조직에 맞게 확장.

- ✓ 관리자·스튜어드 교육
- ✓ 커넥터·플러그인 개발
- ✓ 사내 시스템 연동 개발

☎ 문의: 유지관 · 010-2356-5921 · trendy.ryu@data-dynamics.io

NEXT STEPS

결론 및 다음 단계

Argus Catalog는 데이터·모델·API·AI 에이전트를 하나의 카탈로그에서 거버넌스하는 오픈소스 메타데이터 플랫폼입니다.

가벼운 인프라로 시작해, 표준·품질·리니지·페더레이션·에어갭까지 — 조직의 거버넌스 성숙도를 따라 함께 자랍니다.

1

요건 정의 · 데모

관리 대상 자산·소스 인벤토리, 표준화/망분리 요건 정의, 표준 데모 시연

2

PoC — 실제 소스로 검증

대표 소스 연동 → 동기화·리니지·품질·검색을 실데이터로 확인

3

단계적 확산

파일럿 도메인 → 표준·품질 정착 → 전사 확산·페더레이션 연함

Data Dynamics Inc

문의: 유지관 · 010-2356-5921 · trendy.ryu@data-dynamics.io · www.data-dynamics.io